DOCUMENT RESUME

ED 054 217 TM 000 792

AUTHOR Pascale, Pietro J.; Jakubovic, Shaena
TITLE The Impossible Dream: A Culture-Free Test.
PUB DATE [71]
NOTE 25p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Compensatory Education, Cultural Differences, Cultural Factors, *Culture Free Tests, *Disadvantaged Youth, *Intelligence Tests, Nonverbal Tests, Performance Criteria, *Test Bias, Testing, *Testing Problems, Test Interpretation, Test Wiseness

ABSTRACT

The study reviewed the formats and psychometric rationale of several alleged culture-fair tests. Advantages and disadvantages of each instrument were examined and implications for compensatory education were discussed. (Author)

ED0 54217

THE IMPOSSIBLE DREAM:

A CULTURE-FREE TEST

PIETRO J. PASCALE

and

SHAENA JAKUBOVIC

GRADUATE SCHOOL OF EDUCATION
RUTGERS - THE STATE UNIVERSITY

U.S. DEPARTMENT OF HEALTH,
EDUCATION,& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TM 000 292

ABSTRACT

The study reviewed the formats and psychometric rationale of several alleged culture-fair tests. Advantages and disadvantages of each instrument were examined and implications for compensatory education were discussed.

THE IMPOSSIBLE DREAM: A CULTURE-FREE TEST

Although tests have been in existence informally for millenia, the quest for a culture-free test has only been in the offing since 1926. When it was apparent that subjects from the lower socioeconomic strata consistently scored poorly on the conventional type of intelligence test, e.g., Stanford-Binet (SB), an attempt was made to get at this nebulous "intelligence" sans the frills of culture as if culture were the culprit causing the attainment of low scores on intelligence tests.

Actually the first formal indication that intelligence testing was about to begin occurred in 1896 when Galton conducted his study of genius. He found that men of noted intellect in Britain emanated from just a few families within the country. He attributed the trend of the data to heredity, for he neglected to distinguish one of the intervening variables in intelligence, an enriched environment, which these families offered their members (Barclay, 1968).

The genesis of the culture-free test occurred in 1926 when Davey discovered that pictorial "tests of intelligence" involved G which Spearman held to be the very essence of intelligence, analytical ability. Spearman himself constructed a Visual Perception Test which even

eliminated verbal instructions by using pantomime. He found that such perceptual tests were highly saturated with the G factor. In 1938 Raven's Progressive Matrices appeared, and two years later Cattell's Culture Fair Tests evolved (Cattell, 1940). Other tests which are generally billed as culture free are the Porteus Maze, Goodenough-Harris Drawing Test, Lietner International Performance Scale, Davis-Eells Games, and the non-verbal sections of such tests as the Lorge-Thorndike.

A Look at the Alleged Culture Free Tests

Raven's Progressive Matrices (PM) measures the ability to perceive relationships. The test can be given to children from five and a half years old to adulthood, and the test can be given either individually or in groups. The task is simply to select the design that completes the pattern, so directions can even be given in pantomime. The test is so flexible that one may speed the test or not depending upon the situation. The test can be given on a formboard for very young children, and the task is thereby simplified, for he need only choose a block and put it in the blank. Many tests use this format, but not the same items since there is the chance that subjects may have taken the test before and learned the items. Because of the many arbitrary administration procedures and general lack of standardization, one cannot compare results from various administrations.

The virtues of the PM are that it uses concepts familiar to most, and it is less dependent on education than most intelligence tests. Its greatest fault is that it samples a narrow group of abilities, so one must use some type of vocabulary test to augment its lack of knowledge concerning verbal ability. Although it may measure pure G, it cannot predict as well as a composite of pure G plus verbal, spatial, and other type tests which may give an indication of the skills needed for the job or course for which one wants to predict. The matrix test does show the ability of the subject to direct attention to information, process it, and regulate thought even in those who have not developed reading and verbal skills fully. The empirical evidence shows that those who have not had a great deal of education generally have lower scores, however, than those with more education. Barrett's form of 1956 was found to correlate .75 with the full-scale WISC (Cronbach, 1970). In a study of 30 bilingual Hispanamerican students (8-13 years old), the Raven Colored Progressive Matrices appeared to be the best predictor of school success among other non-verbal and verbal intelligence tests. It was hypothesized that perceptual-motor skills were being used by these children in lieu of verbal skills (Philippus, 1967).

The Lorge-Thorndike (LT) has five verbal and three non-verbal subtests each of which have time limits all totaling 62 minutes plus time for instruction. Although it

has time limits, it has been found that when such were not imposed extra points were not added to the scores. The test is constructed in five levels so that a subject is not given impossibly difficult items nor too easy ones. There are two separate booklets: One for kindergarten through second grade and the other for grades three through twelve. The non-verbal subtests consist of classifications, analogies, and number series which together bear close resemblance to Spearman's G. The correlation between the non-verbal section and the PM is .63, and between it and the WISC Performance the correlation coefficient is .70. The SAT mathematical score and the non-verbal section correlate .70, and furthermore Cronbach (1970) sees this part of the LT as being equal to Cattell's Culture Fair Test in every respect. However, it is interesting to note that the manual advises one to administer lower levels to lower socioeconomic groups than to higher socioeconomic groups. This seems to imply some cultural factor operating here.

Cattell's Culture Fair Test (IPAT CF) has three scales, the first of which is designed for six to eight year olds, the second for ten to fifteen year olds, and third for seventeen to eighteen year olds. The tests' various subtests supposedly tap both crystallized and fluid general abilities of which Cattell and Butcher (1968) say that crystallized ability depends on culture, but fluid ability shows adaptability to new situations where crystallized skills are not applicable. Spearman's G seems to approximate Cattell's

fluid ability.

Piaget's distinction between instruction and construction seems to parallel the distinction between crystallized and fluid ability. Jensen likewise partitions learning into two constructs. Level 1 and Level 2 types of learning abilities approximates Piaget's instruction-construction scheme. Level 1 involves the retention of input and the productive capacity of repetition. Level 1 has been called associative learning and Level 2 has been called conceptual or problem solving ability. Level 2 learning involves the manipulation and transformation of material which sounds very much like the assimilation-accommodation process which Piaget has termed construction.

Scale 1 of the IPAT consists of substitution, classification, mazes, selecting named objects, following directions, wrong pictures, riddles, and similarities. In various combinations, these subtests can be constructed into tailor-made tests for individual or group administration, or to get scores specifically for fluid or crystallized ability. There are two equivalent forms, each of which take 35 minutes to administer completely. Scale II takes 25 minutes for administration and 25 minutes for instructions, and there are also two equivalent forms for this scale. There are four subtests --series, classifications, matrices, and topology--which supposedly tap fluid general ability. A special test was devised to be used to supplement Scale II in an effort to tap crystallized general ability. Scale III has the same format

as Scale II, but qualitatively more difficult items, and can be used in conjunction with the Cattell-Scale III as a predictor of graduate school success. Cattell attempted to use content in his tests which is universal in nature and "overlearned," e.g., parts of the human body. He opposes strictly performance tests on the grounds that they avoid knowledge and verbal skills thereby losing knowledge itself (Cattell, 1940).

Cattell found that the asymptotic level of the IPAT CF coincided with that of biological maturation, whereas the asymptotic level of the traditional intelligence tests seem to parallel the age when formal schooling terminates. He further maintains that the norms remain more constant than those of the conventional tests which he advocates should be restandardized every two or three years in an effort to keep up with cultural changes. The traditional tests have been found to be more accurate than the IPAT CF over the short term, but the IPAT CF is more accurate over the long run. Although the fluid ability is considered of fundamental importance, it is remarked that both fluid and crystallized abilities are necessary for efficient intellectual performance, and "In real life, it will also depend to some extent on personality and motivation factors..." (Cattell & Butcher, 1968, p. 21).

Dickinson (1968) used the IPAT CF in an experiment with first graders in which he had an experimental group trained in the skills of classifying things on five different

levels of sophistication over a period of eight months. The control group went up in achievement as measured by the Stanford Achievement Test, but the experimental group scored significantly higher on the IPAT CF and on the California Short-Form Test of Mental Maturity (full scale). This type of data seems to be a serious challenge to the idea of fluid general ability being more fundamental than crystallized since one can teach fluid ability as well; it just seems to be a different area of knowledge that has to be taught--process rather than content.

The Goodenough-Harris Drawing Test asks that the subject draw the best man and woman that he is able to. The test covers the age ranges of preschool to fifteen years of age, and Cronbach (1970) suggests that it be used to supplement the WISC and SB, but not as a substitute. By virtue of the format, the test is simple to administer and scoring rules have been carefully prepared. This test has been found to be bound with some cultural influences for very divergent cultures, e.g., Moslem Arabs have a taboo regarding drawing of images, and the Hopi men produce the ceremonial art and thereby have IQs one standard deviation above those of the women (Hunt, 1969). Some subcultures do not allow for as much opportunity to draw as others, and in conjunction with this fact separate Negro norms are provided for more discriminating evaluations.

In an experiment using first graders from Anglo-Saxon and Spanish backgrounds, the Drawing Test, the LT Form A,

and the California Achievement Test (CAT) Form W were administered in an effort to locate a test which minimized cultural bias. It was found that the Drawing Test and the LO were nearly identical in predicting the CAT, and that furthermore the Drawing Test tended to bring the two groups closer together in the IQ distribution (Schroeder & Bemis, 1969). Anastasi and deJesus (1953) found that mean IQs of Puerto Rican nursery school children in New York City were not significantly different from mean scores previously established for white and Negro children from the same neighborhood as measured by the Drawing Test. The empirical evidence speaks well for the Drawing Test, but one must recall Cronbach's caveat to use it as a supplement and not as a substitute for the conventional intelligence tests.

The Davis-Eells Games consist of pictures with oral instructions given by the examiner. The contents consist of everyday experiences of children, which include probability, money, "best ways," and analogies, and the vocabulary is supposedly common to all urban American children. The items were selected on the basis of reasonableness of problems as indicators of general problem-solving ability. The test is not speeded, for it was felt that speed is culturally contaminated since it tests quick recall vis-à-vis the ability to solve problems. The examiner is instructed to show warmth and encouragement to the subjects. There are two forms: The Primary Test for grades one and two and the Elementary Test for grades three through six (Ahmann, Glock, & Wardeberg,

1960). Coleman and Ward (1955) compared the Games and the Khalmann-Finch scores of children from low and high socio-economic groups, and they found no significant differences between the two. Ludlow (1966) found that the Games were not superior to conventional test means, and furthermore that lower-class retarded children had significantly higher scores on the WISC Performance than on the Games. These children showed no significant differences on their scores between the Games and the SB, WISC, or the California Test of Mental Maturity (CMM). It seems clear from these data that the Davis-Eells Games is not immune from culture as its authors would have us believe.

Arguments: Pro and Con

Now that we have looked at the format of some of the "culture-free" tests and some of the empirical evidence, it is appropriate to investigate the pros and cons of such instruments. The first argument involves an indictment against all intelligence tests in general. Great fluctuations are found in IQ scores for children six to eighteen years of age. Data show the IQ is not a reliable concept since the test which measures it, the grade level at which it is measured, and the norms by which it is evaluated can make a difference in "IQ". It has been found "that changes in mental test scores tend to be in the direction of the family level, as judged by the parents' education and socio-economic status" (Honzik, Macfarland, & Allen, 1966, p. 172). It has been estimated that a score on a six-year test could

change 20 points for one out of three children by the age of 18, and 15 points for six out of ten children, thus seeming to invalidate the predictive validity of an intelligence test score which one is prone to take at face value (Honzik, et al., 1966). One basic violation that constantly takes place in evaluation is that testing provides an objective basis for assessment and evaluation of an individual's characteristics whereas in reality such testing is most valid for groups, and least valid for individuals. Tests can only give qualified information which is dependent on many possible sources of error for individuals. In all cases test scores should be analyzed in terms of the cultural grouping's overall performance, and the criterion of excellence as purported by the "average" should not be ideal. "The mathematical symbol becomes all too often the criterion of expected performance without any real reference to the behavioral phenomena needed for success in the cultural setting" (Barclay, 1968, p. 26).

Vernon (1965) states that the cultural level of the home is the single most significant influence on the scores (even above SES) of intelligence tests. If this is true, and it is also true that a four-year-old's IQ correlates about .70 with late adolescence's (Cronbach, 1970), then equal education can be of very little help since the lower-class children are not adequately prepared for school, and the upper-class children are more than adequately prepared to assume the role of student. Strodtbeck (1965) has isolated some of the factors which create this situation. He

advocates the position that lower-class chi'dren are doomed to failure, because they do not develop test-taking skills, responsiveness to speed requirements, and familiarity with vocabulary. He feels that these are the very reasons why the lower-class child shows poor performance in the classroom, on the job, as well as on tests. The question is posed if the importance of verbal intelligence should be diminished in the school environment, or if the curriculum should be altered to develop verbal intelligence throughout the entire population. This inquiry leaves the concept of culture-free tests in a very tenuous position.

Eells, et al. (1951) define other characteristics of the middle-class home which the lower-class home lacks. The child in the middle-class home learns to make a good impression, and to develop certain attitudes toward himself, and toward task performance. These influence his response to tests and to school assignments which he learns to take seriously since he is constantly given tangible and intangible reinforcement, whereas the lower-class child learns to take his assignments lightly and just work to keep out of trouble. His rewards are the intangle ones which he gets from his peers and not from approving adults.

Hertzig, et al. (1968) found similar evidence in a study of Puerto Rican preschoolers. Children in Puerto Rican homes find their reinforcement contingent on effective social interactions rather than on task mastery, which is the basis of our task-oriented schools. Since Puerto Rican and middle-class SES children differ at three years of age from each

other in their styles of response to demands for cognitive functioning, the continuation of this manifested in the school situation is practically assured and so is their failure in United States' schools. Although differences in behavioral style may have differential consequences, the presence of a difference should not be construed to mean that one pattern is superior or inferior to another. How effectively would a middle-class white student function in a South American school? Just as Hertzig et al. found a difference in behavioral styles of different ethnic groups, so did Lesser, Fifer, and Clark (1965) find that there are differences in the hierarchical patterns of intellectual skills between ethnic groups but not within them. The ramifications of this data belie issue of how culture-free test authors can presuppose common experiences if there may not be any at all in fact.

There is some psychophysical data that suggest that some conceptual difficulties may originate in perceptual insensitivities (Farnham-Diggory, 1970). Perceptual difficulties have been found in disadvantaged children which makes it a problem for them to find similarities, differences, and relations of the part to the whole (Klaus & Gray, 1968). This type of evidence would make one wonder exactly how much can such a child discern in a test which relies purely on perceptual abilities, in fact, is known as a perceptual intelligence test. If there were definitive evidence forthcoming regarding perceptual inadequacies, there would be two major types of research--why does it occur (organically based

or lack of experiential opportunities) and how to overcome these known difficulties through alternative routes which could compensate for these difficulties.

Lorge (1945) supports the notion that education does make an inherent difference on intelligence tests. He found that those with greater education scored significantly higher than those with less education. The evidence comes from re-tests of equated groups of subjects after twenty years. Education increases the mastery of abilities measured by IQ tests while lack of further education diminishes the mastery of such abilities. His argument is that superior intellectual ability needs stimulation, and that full potentials may be lost in the absence of such stimulation. Along the same vein, Brazziel (1969, p. 207) says, "If conceptual learning is viewed as a gradual acculturation process and offered early in school careers, these children (inner city blacks) can be made to think."

As if these complications would not be enough for the test writers to cope with, the mere idea of a strange examiner can often be found to be instrumental in depressing scores. It may be advisable in some cases to spend some time with the subject until he is at ease with the examiner, or an examiner with the same background may be indicated. Abstract-type tasks may seem of little value for some to bother with since their culture may involve itself only with things of practical significance. If a subject does not have a familiarity with pictures and diagrams, or even paper and

pencil, scores which are reflective of the subject's ability may not be elicited (Anastasi, 1960).

After this brief glance at some of the intervening variables which cause low scores on intelligence tests, one can understand why the search for a culture-free instrument would be desirable. However, one must realize that the possibility exists that an intelligence test cannot be constructed which does not discriminate among classes since social class differences are real. Stroud (1957, p. STD-9A) issues an injunction against the concept of culture-free tests when he says that "...the cultural impact associated with social class differences may affect the course of mental development of children as well as their performance on intelligence tests." Bradfield and Moredock (1957, p. 375) give added credence to Stroud's thoughts when they say, "Speaking and reading, writing and listening, of all human behaviors, perhaps necessitates the most discrimination, memory, generalization, etc., and hence you should expect the children more skilled verbally to be more intelligent." This may not be as far fetched as it first sounds, for if it is true that language is a tool for the development of the intellect as Piaget (1967) would have us believe, then those who do not have formal linguistic skills would not be as well equipped in the area of intellectual ability. Unfortunately formal linguistic skills taught at school in a foreign language (even blacks use another language at home) are not always flexible in nature, and therefore, they do not become part of

the intentional repertoire of these marginal children.

Anastasi (1958a, p. 534) questions the feasibility of constructing a culture-free instrument on the grounds that perhaps the concept of intelligence is itself culturally conditioned and restricted. "It is not so much that tests are unfair to lower-status groups, as that lower-class environment is not conducive to the effective development of 'intelligence' as defined in our culture." Lesser, et al. (1965, p. 12) quotes Lorge:

> There is no virtue in developing instruments so blunted that they decrease the amount of information. Perhaps the best method of reducing bias in tests of intelligence is to use them with the full knowledge that endowment interacting with opportunity produces a wide range of differences. Appraisal of the variation of different kinds of intellectual functioning requires many kinds of tests so that the differences can be utilized for the benefit of the individual and for the good of society. Intellectual functioning certainly does involve the ability to learn to adjust to the environment or to adapt the environment to individual needs and capabilities by the process of solving problems either directly or incidentally. Such a concept recognizes a variety of different kinds of problems. The full appreciation of the variety of aptitudes and the development of adequate methods for appraising them, should in the long run, ultimately lead to the production of enough information to eliminate bias.

To put it more succinctly, "...to rule out cultural differentials from test items so as to make them equally 'fair' to subjects in different social classes or in different cultures may merely limit the usefulness of the test, since the same cultural differentials may operate within the broader area of behavior which the test is designed to sample." (Anastasi, 1958b, p. 202).

Culture-free intelligence tests may go on being a controversial issue in spite of all the logical arguments which oppose such instruments, but the fact that culture-sterile materials are only a stopgap measure should become apparent to all. The immediate suggestion for the long-range outlook would appear to be compensatory education, but one might ask in what language and whose curriculum. It seems valid to suggest that children from backgrounds with different languages at home be taught bilingually until they are able to switch to either with equal facility. Culture infested tests of intelligence would then give a true picture of the progress of process and not of the stagnation which it now records.

The most prolific authority on the ontogenesis of intelligence, Piaget (1967), posits that humans are endowed with two functional invariants, i.e., adaptation and organization--two modus operandi, but that experience and environment intervene to activate the mechanism, and if the unit is not frequently and well nourished, the mechanism will never develop its full potential. Anastasi (1960) notes that changes in intelligence test scores are influenced by intervening experiences which run the gamut from emotional to environmental experiences. In general, the underprivileged environment rears children who lose IQ points with age, and superior environments rear children who gain IQ points. Psychologists have found significant increases in mean scores which coincide with socioeconomic and educational improvements.

Pepin (1971) quotes Bruner as saying that the objective of teaching should be to discover the limits of capacities, i.e., how far can the child go given the best instruction, and that he advocates teaching and testing alternately to show where the child is capable of going. The reality of social conditions must be corrected and not a score. The idea is to find out the educational diagnosis and then find a cure for the deficiencies rather than predicting the death of the intellect of the patient. Poor performance on IQ tests should be incentives and challenges to educators who should endeavor to modify scores in an upward direction. There is no damage done to the validity of a test if one coaches a student in making better test scores if such coaching is instrumental in improving a general area of intellectual skills as well as would be accomplished through a carefully constructed curriculum. Just as content can be taught, so can process.

Piaget (1970, p. 714) says that "...learning is no more than a sequence of cognitive development which is facilitated or accelerated by experience." He contends that teaching through environmental activities and experiences can accelerate or complete structures which are being developed but that the order to succession will remain invariant. He is of the opinion that the optimal situation for learning is one in which the child discovers the new information himself (could this be the needed curriculum?), but he does advocate teachers to devise situations in which there would be the opportunity for such unfolding to occur. He contends that exercise,

experience, and social action are the variants between heredity and actualization. The educational implications are very grave ones, and it is not in the interest of children from any social class to leave them go unheeded by alleviating the problem with a correction factor on a test.

Attempts at Operational Definition

The fact of mean group differences does not *per se* indicate that a test is unfair or biased for the group with the lower mean. Cronbach (1970) suggests that regression analysis (not t-testing) procedures are appropriate for examining and determining test bias. Test bias has been defined by Cleary (1968, p. 115):

> A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test is designed, consistent nonzero errors of prediction are made for members of the subgroups. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroups.

As Linn and Werts (1971) point out, the critical component of the Cleary definition is that the criterion variable be free of bias. Rarely do we find in educational and psychological data criteria completely bias-free. The problem is further impacted by the unreliability of the criterion variable.

Determination of test bias is clearly a function of test use and interpretation. Scores *in se* do not determine test bias (Thorndike, 1971). We may say that test bias is situation-specific. A final note is the issue of criterion relevance. Frequently, a paper and pencil test used as the

criterion has little, if any, relationship to the tasks required for job success. In this case it is inappropriate to label the test unfair. Its use should be termed irrelevant (Thorndike, 1971).

REFERENCES

Ahmann, J. S., Glock, M. D., & Wardberg, H. L. Evaluating elementary school pupils. Boston: Allyn and Bacon, Inc., 1960.

Anastasi, A. Differential psychology. New York: The Macmillan Company, 1958a.

Anastasi, A. Heredity, environment, and the question "how." Psychological Review, 1958b, 65, 197-208.

Anastasi, A. Standardized ability testing. In P. H. Mussen (Ed.) Handbook of research methods in child development. New York: John Wiley & Sons, Inc., 1960, Pp. 456-486.

Anastasi, A., & deJesus, C. Language development and Good-enough Draw-a-Man I.Q. of Puerto Rican preschool children in New York City. Journal of Abnormal Social Psychology, 1953, 48, 357-366.

Barclay, J. R. Controversial issues in testing. Boston: Houghton Mifflin Company, 1968.

Bradfield, J. M., & Moredock, H. S. Measurement and evaluation in education. New York: The Macmillan Company, 1957, Pp.

Brazziel, W. F. A letter from the South. Harvard Educational Review, 1969, 39 (2), 200-208.

Cattell, R. B. A culture free intelligence test: I. Journal of Educational Psychology, 1940, 31, 161-179.

Cattell, R. B., & Butcher, H. J. The prediction of achievement and creativity. Indianapolis: The Bobbs-Merrill Company, Inc., 1968.

Cleary, T. A. Test bias: prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Coleman, W., & Ward A. Comparison of Davis-Eells and Kuhlmann-Finch scores of children from high and low socio-economic status. Journal of Educational Psychology, 1955, 46, 465-469.

Cronbach, L. J. Essentials of psychological testing. (3rd ed.) New York: Harper & Row, Publishers, 1970.

Darlington, R. D. Another look at "cultural fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Dickinson, D. J. The effects of practice in making classifications on achievement and intelligence. Unpublished paper, 1968.

Eells, K. W., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. W. Intelligence and cultural differences. Chicago: University of Chicago Press, 1951.

Farnham-Diggory, Sylvia. Cognitive synthesis in Negro and white children. Monographs of the Society for Research in Child Development, 1970, 35 (2).

Hertzig, M. E., Birch, H. G., Thomas, A., & Mendez, O. A. Class and ethnic differences in the responsiveness of preschool children to cognitive demands. Monographs of the Society for Research in Child Development, 1968, 33 (1).

Honzik, M. P., Macfarland, J. W., & Allen, L. The stability of mental test performance between two and eighteen years. In C. I. Chase & G. H. Ludlow (Eds.) Readings in Educational and Psychological Measurement. Boston: Houghton Mifflin Company, 1966, Pp. 153-172.

Hunt, J. McV. Has compensatory education failed? Harvard Educational Review, 1969, 39 (2), 130-152.

Klaus, R. A., & Gray, S. The early training project for disadvantaged children: A report after five years. Monographs of the Society for Research in Child Development, 1968, 33 (4).

Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. Monographs of the Society for Research in Child Development, 1965, 30 (4).

Linn, R. L., & Werts, C. E. Considerations for studies of test bias. Journal of Educational Measurement, 1971, 8, 1-4.

Lorge, I. Schooling makes a difference. Teachers College Record, 1945, 46, 483-492.

Ludlow, G. H. Some recent research on the Davis-Eells Games. In C. I. Chase & G. H. Ludlow (Eds.) Readings in Educational and Psychological Measurement. Boston: Houghton Mifflin Company, 1966, Pp. 206-210.

Pepin, A. C. The IQ test: Education's bugaboo. The Clearing House, 1971, 45 (5), 278-280.

Philippus, M. J. Test prediction of school success of bilingual Hispanamerican children. Unpublished paper, 1967.

Piaget, J. Six psychological studies. New York: Vintage Books, 1967.

Piaget, J. Piaget on his theory. In P. Mussen (Ed.) Carmichael's manual of child psychology. New York: John Wiley, 1970, Pp. 703-732.

Schroeder, G. B., & Bemis, K. A. The use of the Goodenough Draw-a-Man Test as a predictor of academic achievement. Unpublished paper, 1969.

Schwebel, Milton. Who can be educated? New York: Grove Press, 1968.

Strodtbeck, F. L. The hidden curriculum in the middle-class home. In J. D. Krumboltz (Ed.) Learning and the Educational Process. Chicago: Rand McNally & Company, 1965, Pp. 91-112.

Stroud, J. B. The intelligence test in school use: Some persistent issues. In P. J. Pascale (Ed.) Basic readings in tests and measurement. New York: Simon & Schuster, Inc., 1970, Pp. STD-1A-9A.

Thorndike, R. L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Vernon, P. E. Ability factors and environmental influences. American Psychologist, 1965, 20, 723-733.